



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2020

Radiotherapy Treatment plannINg study Guidelines (RATING): A framework for setting up and reporting on scientific treatment planning studies

Hansen, Christian Rønn ; Crijns, Wouter ; Hussein, Mohammad ; Rossi, Linda ; Gallego, Pedro ;
Verbakel, Wilko ; Unkelbach, Jan ; Thwaites, David ; Heijmen, Ben

Abstract: Radiotherapy treatment planning studies contribute significantly to advances and improvements in radiation treatment of cancer patients. They are a pivotal step to support and facilitate the introduction of novel techniques into clinical practice, or as a first step before clinical trials can be carried out. There have been numerous examples published in the literature that demonstrated the feasibility of such techniques as IMRT, VMAT, IMPT, or that compared different treatment methods (e.g. non-coplanar vs coplanar treatment), or investigated planning approaches (e.g. automated planning). However, for a planning study to generate trustworthy new knowledge and give confidence in applying its findings, then its design, execution and reporting all need to meet high scientific standards. This paper provides a 'quality framework' of recommendations and guidelines that can contribute to the quality of planning studies and resulting publications. Throughout the text, questions are posed and, if applicable to a specific study and if met, they can be answered positively in the provided 'RATING' score sheet. A normalised weighted-sum score can then be calculated from the answers as a quality indicator. The score sheet can also be used to suggest how the quality might be improved, e.g. by focussing on questions with high weight, or by encouraging consideration of aspects given insufficient attention. Whilst the overall aim of this framework and scoring system is to improve the scientific quality of treatment planning studies and papers, it might also be used by reviewers and journal editors to help to evaluate scientific manuscripts reporting planning studies.

DOI: <https://doi.org/10.1016/j.radonc.2020.09.033>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-193287>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.

Originally published at:

Hansen, Christian Rønn; Crijns, Wouter; Hussein, Mohammad; Rossi, Linda; Gallego, Pedro; Verbakel, Wilko; Unkelbach, Jan; Thwaites, David; Heijmen, Ben (2020). Radiotherapy Treatment plannINg study

Guidelines (RATING): A framework for setting up and reporting on scientific treatment planning studies.
Radiotherapy and Oncology, 153:67-78.
DOI: <https://doi.org/10.1016/j.radonc.2020.09.033>



Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Review Article

Radiotherapy Treatment planning study Guidelines (RATING): A framework for setting up and reporting on scientific treatment planning studies

Christian Rønn Hansen^{a,b,c,d,*}, Wouter Crijns^{e,f}, Mohammad Hussein^g, Linda Rossi^h, Pedro Gallegoⁱ, Wilko Verbakel^j, Jan Unkelbach^k, David Thwaites^c, Ben Heijmen^h

^a Laboratory of Radiation Physics, Odense University Hospital; ^b Institute of Clinical Research, University of Southern Denmark, Odense, Denmark; ^c Institute of Medical Physics, School of Physics, University of Sydney, Sydney, Australia; ^d Danish Centre for Particle Therapy, Aarhus University Hospital, Denmark; ^e Department Oncology - Laboratory of Experimental Radiotherapy, KU Leuven; ^f Radiation Oncology, UZ Leuven, Belgium; ^g Metrology for Medical Physics Centre, National Physical Laboratory, Teddington, UK; ^h Erasmus MC Cancer Institute, Radiation Oncology, Rotterdam, The Netherlands; ⁱ Servei de Radiofísica i Radioprotecció, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain; ^j Amsterdam University Medical Center, The Netherlands; ^k Radiation Oncology Department, University Hospital Zurich, Switzerland

ARTICLE INFO

Article history:

Received 1 July 2020

Received in revised form 4 September 2020

Accepted 11 September 2020

Available online xxxx

Keywords:

Guidelines

Radiotherapy

Treatment planning

Plan comparison

Plan quality

ABSTRACT

Radiotherapy treatment planning studies contribute significantly to advances and improvements in radiation treatment of cancer patients. They are a pivotal step to support and facilitate the introduction of novel techniques into clinical practice, or as a first step before clinical trials can be carried out. There have been numerous examples published in the literature that demonstrated the feasibility of such techniques as IMRT, VMAT, IMPT, or that compared different treatment methods (e.g. non-coplanar vs coplanar treatment), or investigated planning approaches (e.g. automated planning). However, for a planning study to generate trustworthy new knowledge and give confidence in applying its findings, then its design, execution and reporting all need to meet high scientific standards. This paper provides a 'quality framework' of recommendations and guidelines that can contribute to the quality of planning studies and resulting publications. Throughout the text, questions are posed and, if applicable to a specific study and if met, they can be answered positively in the provided 'RATING' score sheet. A normalised weighted-sum score can then be calculated from the answers as a quality indicator. The score sheet can also be used to suggest how the quality might be improved, e.g. by focussing on questions with high weight, or by encouraging consideration of aspects given insufficient attention. Whilst the overall aim of this framework and scoring system is to improve the scientific quality of treatment planning studies and papers, it might also be used by reviewers and journal editors to help to evaluate scientific manuscripts reporting planning studies.

© 2020 The Author(s). Published by Elsevier Ireland Ltd. All rights reserved. Radiotherapy and Oncology xxx (2020) xxx–xxx

Motivation and aim of the RATING framework

Treatment planning has advanced substantially since the introduction of the first computerised treatment planning systems (TPS) [1,2]. The evolution of computer hard- and software allowed implementation of advanced algorithms for dose calculation and optimisation, and facilitated the introduction of many complex treatment techniques [3–6]. Radiotherapy treatment planning is currently in an exciting era with novel developments in, for example, automated planning, robust planning, online adaptive radiotherapy, knowledge-based planning and plan quality assessment methods [7–11].

Treatment planning studies serve several purposes. They play an important role in developing, verifying and implementing advanced treatment techniques in the clinic [12]. For comparing treatment techniques, they act as *in silico* surrogates for clinical studies, particularly where these are considered not feasible, leaving planning studies as an alternative to investigate the added value of new approaches. They can also have more technical aims, e.g. in the development of improved optimisation algorithms and methods, or the evaluation and comparison of commercial TPS. Whatever the application, studies should be carefully and robustly designed and conducted. Also, the work should be reported with sufficient and consistent detail to ensure: it can be reproduced, the findings can be evaluated with confidence by others, and the methods can be safely and reliably applied elsewhere [12].

More generally, treatment planning studies are examples of computer simulation, which is increasingly popular in many fields

* Corresponding author at: Odense University Hospital, Sdr Boulevard 29, 5000 Odense C, Denmark.

E-mail address: christian.roenn@rsyd.dk (C.R. Hansen).

<https://doi.org/10.1016/j.radonc.2020.09.033>

0167-8140/© 2020 The Author(s). Published by Elsevier Ireland Ltd. All rights reserved.

[13]. In most applications, the simulations generated require quality assurance, via an informal or formal review process that assesses their credibility or maturity [13], and whether the comparison with other simulations, i.e. in this context comparative plans, is fair. This can be accomplished using credit scores indicating their trustworthiness. This work aims to enhance treatment planning study quality by presenting a maturity assessment framework [14] including a numerical scoring (RATING). Background on the development of this framework is presented in the [supplementary material](#).

The proposed guidelines are intended to assist investigators to design, organise and evaluate the quality of a treatment planning study. They are also aimed at encouraging good practice when reporting planning studies and could potentially be used by reviewers and editors to assess consistently the maturity level of submitted study papers. The guidelines mostly relate to methodological aspects of planning studies, i.e. not to other parameters, e.g. novelty, urgency, relevance of the reported work that might also determine their potential impact and acceptability to a specific journal.

Whilst the focus of the guidelines is on planning studies to answer novel research questions for dissemination to the wider community, the recommendations can also enhance quality and reliability of internal studies, e.g. on the safe implementation of a new treatment technique in the department. To support good practice in dissemination and reporting, the guidelines are given following the general structure of a scientific manuscript, i.e. starting with a description of the planning study design and development, through to the critical discussion of results and conclusion. Each section addresses issues to be considered in both developing and reporting a planning study by providing questions for the researchers to be answered by yes or no, each with a relative points score. The questions are collected in a checklist. Not all questions are relevant for all study types, hence such questions can be selected as applicable/relevant or not and then answered. The RATING score for the study is then calculated as the normalised weighted-sum score (as a percentage) of all applicable questions and can be used to indicate studies carried out and reported within this proposed high quality framework.

Considerations for designing and developing a planning study: the introduction section

For designing and developing a study, there should be a clear reasoning for deciding what is to be studied and why. This will be reflected in the Introduction section of the final study report, which should contain various essential parts, including the background to the study, to help understand its basis and rationale. The introduction must be concise and end with an unambiguous study aim defined by research questions which can be concluded upon at the end of the work.

Given the wide variety of planning study types and scope, formulating the study's aim should also provide an initial understanding of the type of work presented. An example may be the comparison of a commercial automated planning solution to a manual planning technique. Does the study address a scientific question that is independent of a particular TPS, e.g. quantifying the dosimetric differences between proton and photon plans? Does the work contribute to the further development of treatment planning methodology not yet available in commercial systems, e.g. in-house algorithm development that is demonstrated for selected patients?

The study aim formulated by research questions

Any scientific study must have a precisely and concisely defined aim. Generally, there is a primary research question, possibly supported by other secondary questions. The research questions should be limited in number and clearly interconnected and related to the overall study aim. Research questions should be defined at the beginning of the formal study; often preceded by exploratory pilot studies, so appropriate measures can be taken to ensure a systematic and robust study setup.

1. Does the study have a concise and precise study aim, defined with a restricted number of interconnected questions? (10 points, mandatory)

The motivation for the research questions

There must be a clinical, technical or scientific need for the posed research questions, which should be clearly substantiated in the Introduction section. A thorough literature survey is required to exclude duplication of already published work and to identify previously published work that supports the research questions, pinpointing gaps in current knowledge and understanding, and demonstrating the need for further investigation. This may include studies that have investigated similar or the same research questions, e.g. using different scientific approaches. All such relevant work should be concisely discussed and referenced.

2. Has relevant up to date literature been included to support the need for the current study? (5 points, mandatory)
3. Does the study address an existing knowledge gap? (10 points, mandatory)

Considerations for the methodology: the Materials and methods section

Two main aspects should be considered regarding the applied methodology: firstly, it is crucial that the global study design and applied methodology will indeed result in answers to the posed research questions; secondly, in reporting there should be sufficient information for independent researchers to be able to interpret results and reproduce the study in their institutes. In a manuscript, this will be reflected in the Materials and Methods section, which should clearly and unambiguously describe how the work has been done. The following subsections give guidance on considerations necessary when developing the methodology and reporting it.

4. Is the global study design adequate for answering the posed research questions? (10 points, mandatory)
5. Is the global study design described in sufficient detail for others to interpret and reproduce the results? (5 points, mandatory)

Patient cohort

The patient cohort is often the foundation of the planning study and should be described in sufficient detail to provide an understanding of how generalisable the study is, e.g. to a local patient cohort. The selection criteria used for inclusion and exclusion of patients should be detailed to allow readers to understand if the planning study would fit their cohorts. The number of included patients should be stated and should be justified.

The patient cohort size should be carefully considered to ensure that sufficient patient anatomies and target sizes and locations are represented. The most appropriate method of patient selection depends on the research question. Some studies require a large

number of consecutive or randomly selected patients to answer it adequately, whilst others benefit from a detailed and insightful analysis of a few representative hand-selected patients. For some rare indications, large patient cohorts might not even be available.

As for other studies in medicine, also for treatment planning studies, there may be a need for approval by an institutional review board or ethics committee. A statement saying that approval was obtained, or that a request for approval was not needed, should be included in the manuscript.

6. Are the inclusion and exclusion criteria of the patient cohort described? (1 point)
7. Is the clinical patient information of the cohort presented, including disease type, site(s) and clinical staging? (1 point)
8. Is the included number of patients stated, explained and justified? (1 point)
9. Has there been consideration of the need for ethical and/or legal approval for the study and if needed, is there a statement about this? (5 points)

Imaging procedures

Where relevant for the study, the patient imaging and immobilization methods should be described,

10. Have the scanning parameters been reported in sufficient detail (image modalities, equipment model, slice thickness, voxel size, patient position (e.g. head first, supine, etc.)? (1 point)
11. Has the applied immobilisation equipment been described, (e.g. vendor and type, standard settings, etc.)? (1 point)

Treatment machine and settings

Often, the treatment machine and its settings and calibration need to be described in planning studies, which should include manufacturer and model, energies used, MLCs, etc. An often applied metric is the monitor unit (MU). However, this may have limited value without clear definition; if the reference conditions are not defined there can easily be a 20% difference in MU depending on the calibration procedure.

12. Have the treatment machine and relevant parameters been described with sufficient detail (model, beam energy, MLC, etc.)? (1 point)
13. Have the MU reference conditions been defined? (1 point)

Definition of targets and OARs

If available, published protocols may be useful to establish GTVs (gross tumour volume), CTVs (clinical target volume), OARs (organs at risk), PTVs (planning target volume) and PRVs (planning risk volume) in planning studies to more easily apply results in other departments. In any case, sufficient detail should be provided on how the structures were defined including optimisation help structures, with a focus on their extent and sizes (e.g. mean PTV volume with a range). Depending on the study type, it may also be valuable to describe the roles of all involved personnel (RTT, physicist, oncologist, radiologist, etc.). Details of the contouring tools used can also be beneficial; auto-contouring with or without editing, applied software package including version, independent validation of contours, etc.

GTVs and CTVs

For GTV definition, a description of the applied (multi-modality) imaging is generally needed. For CTV definition around a GTV, the

applied protocol and resulting margins should be summarised. For the definition of elective CTVs, it should be clear what nodal regions are included.

14. Has GTV definition been described in sufficient detail, with references if possible? (1 point)
15. Has CTV definition been described in sufficient detail, with references if possible? (1 point)

PTVs

For PTVs, applied margins should be specified, along with their basis, e.g. literature, national protocol, institutionally derived. If probabilistic/robust planning is used with no PTV definition, then this should be clearly stated.

16. Has the establishment of PTVs (or alternatively robustness settings) been described in sufficient detail? (1 point)
17. Have PTV sizes in the patient cohort been described? (1 point)

OARs and PRVs

Non-trivial aspects of OAR definitions relevant for planning or reporting of NTCP values should be explicitly given, e.g. the contoured length of rectum, esophagus or spinal cord, or contouring of the pharyngeal constrictor muscles if they overlay the GTV, or whether esophagus contouring considers the position in all phases of a 4DCT-scan. Reported NTCP predictions almost always rely on dose metrics directly connected to OAR definitions and should therefore preferably follow the definition used in the NTCP-modelling. If applied, margins around OARs to derive PRVs should be described.

18. Have OAR definitions been described in sufficient detail, with references if possible? (1 point)
19. Have PRV margins been described in sufficient detail, with references if available? (1 point)

Treatment planning system and dose calculation

Depending on the planning study type, relevant TPS information should be provided in sufficient detail. Different TPS versions can impact on the options available and are therefore important for reproducing the results or for clinical implementation. All relevant user settings should be reported since these often are key to a successful implementation, i.e. accelerator, beams, dose grid, optimisation grid, control point spacing etc. If plans from different TPSs are compared in the study, they should preferentially be compared in the same software with a common dose sampling and evaluation. This is especially important when small volumes are evaluated [15].

20. Have all applied dose calculation algorithms been described in sufficient detail? (1 point)
21. For any commercial software used, have the manufacturer, algorithms and specific versions been stated? (1 point)
22. Have all relevant user parameters and settings in the TPS been reported, e.g. beams, dose grid, control point spacing? (1 point)
23. Have all volumes been evaluated with the same software/methodology? (1 point)

Planning aims and optimisation

A clear description of the planning aims and optimisation approach is essential, i.e. how an optimal plan should be and how it was generated. Important to note here is that cost functions

used in the TPS for plan generation are sometimes different from the functions defining the true or clinical planning aims. This may, for example, happen if a cost function used in the planning protocol is not available in the TPS, or if a surrogate convex function is preferred for planning instead of the non-convex function used in the protocol, to avoid getting trapped in a local minimum.

Of major importance is a clear description of the applied dose prescription; dose to the isocentre, to the isodose including 98% of the PTV volume, etc. It is also important to discriminate between hard constraints that render a plan unacceptable if violated, and objectives that should be achieved as well as possible, but without causing a violation of imposed hard constraints. Objectives are generally not equally important, i.e. in plan generation, they have different rankings or priorities, which in many TPS may translate to differences in weights of cost functions. For example, sparing of one OAR may be more important than another, or reduction of small volumes with high dose in an OAR may be more important than reduction of the mean dose. Therefore, both plan acceptability and plan quality should be well described as they drive the optimisation. Differences in priorities can often significantly influence the planning outcome.

Objectives do not always have specified goal values, e.g. for high or low dose conformality, or dose homogeneity in the PTVs (e.g. in SBRT). If there are goal values, they are sometimes called soft constraints. In the latter case, one planning aim is to meet all soft constraints. However, violation of a soft constraint does not necessarily make a plan unacceptable for patient treatment (see below, 'Plan acceptability – minor and major protocol deviations'). Both for hard and soft constraints, it is good practice to try to reduce dose to below the constraint level. Also for this, there may be priorities.

Planning studies need a description of the optimisation approach, including all aspects that influence final dose distributions, including the use of the defined priorities/ranking of all objectives, manual or automated planning, applied optimisation structures, applied number of iterations, etc. For some planning studies, it can be useful if the planning aims and applied optimisation approach align with a published planning protocol (e.g. DAHANCA, RTOG), defining requirements for target and OARs and mutual ranking.

24. Are clear planning aims defined, including imposed hard constraints and planning objectives (with or without soft constraints)? (5 points, mandatory)
25. Has the ranking of planning objectives (priorities) been described? (5 points, mandatory)
26. Is the dose prescription clearly defined? (10 points, mandatory)
27. Is there a description of the applied optimisation process, including the handling of all objectives with their ranking? (5 points, mandatory)
28. If manual intervention during or after optimisation is allowed, has this been described? (1 point)

Bias mitigation

Planning studies comparing treatment or planning approaches are prone to bias, which can result in incorrect answers to the research questions. Whether or not there is bias can depend on the planning study type. If TPS for conventional trial-and-error planning is mutually compared, it is important that the planner is as experienced for each TPS used. In the comparison of automatic vs. manual planning, the retrospective inclusion of recent manual plans to compare with autoplans does not lead to bias if the research aim was to compare autoplanning with routine manual

planning, i.e. comparison with not necessarily the best possible manual planning. If the aim was to compare with the best possible manual planning, then prospective planning by the best planner in the best conditions would be a better alternative, or even better, independent planning by a group of best planners. In treatment technique comparisons, using retrospectively included plans for one technique vs. prospective planning for another is likely to introduce unacceptable bias. In particular, in studies evaluating technical development, it can be problematic to use clinical plans as the baseline, since these potentially have numerous compromises (e.g. related to available planning time), and may therefore not be representative of the true strengths of the conventional technique. If several delivery techniques (e.g. coplanar and non-coplanar treatment) are compared with different TPSs with different levels of sophistication, there is a clear bias in the technique comparison. However, if several TPSs of variable sophistication are used, the planning study may still answer a relevant clinical question, e.g. in case one of the techniques has its own dedicated TPS. Serious bias can also be introduced where alternative plans for patients are generated while having knowledge of the original plans, or if planning times for one group of plans are clearly longer than for a competitive group of plans.

29. Have enough study details been provided such that bias issues could be noted? (5 points, mandatory)
30. Has bias been sufficiently mitigated to reliably answer the posed research question? (10 points, mandatory)

Plan acceptability – minor and major protocol deviations

Plan acceptability, i.e. the suitability of generated dose distributions for clinical treatment, is generally of high importance in a planning study. In the planning protocol, there can be definitions for minor or major protocol deviations (e.g. for a planning study in the context of a formal clinical study). The outcome of plan acceptability evaluations, and of minor or major deviations, are always binary, i.e. yes or no answers. Plan acceptability can be assessed by comparing achieved plan parameters with imposed hard constraints. Depending on the study, plans could also become unacceptable with too many or too large deviations of soft constraints (see, 'Planning aims and optimisation approaches'). Analysis of plan parameters, to assess plan acceptability and minor and major deviations, can be supplemented by a formal evaluation of plans by a clinician for the whole cohort or a subgroup. This is highly recommended for planning studies with rather clinical research questions.

31. Was the procedure for assessment of plan acceptability well described? (1 point)
32. Was the procedure for assessment of minor and major protocol deviations well described? (1 point)

Plan (re-)normalisation for plan comparisons

For various reasons, it can be useful to (re-)normalise generated plans before final evaluations or comparisons. For example, where there are slight variations in PTV coverage between plans, re-normalisation can ensure that coverage is the same for all involved plans. Where other PTV dose characteristics are also similar, further plan evaluations and comparisons can then focus on OAR doses. However, it should be noted that (re-)normalisation can change the optimisation optimum and only small changes are recommended.

33. Has plan (re-)normalisation been described sufficiently? (1 point)

Dose–volume parameters for plan evaluation and comparison

Plan evaluations and comparisons should always include the most important plan parameters used for defining the planning aims (constraints and objectives). As explained in ‘Planning aims and optimisation’, these parameters may not always be the same as the obtained values for the cost functions used for plan generation. Additional parameters may also be used.

34. Have sufficiently comprehensive dose–volume parameters been used for plan evaluations and comparisons? (5 points, mandatory)

Population-mean DVHs

Population-mean or median DVHs may provide added value to planning studies. If used, it is recommended to include confidence intervals to visualise the significance of differences. How population-mean DVHs and confidence intervals are derived should be clearly defined. Statistical comparison of average DVHs has been described by Bertelsen et al [5].

35. Has the algorithm for creating population-mean/median DVHs been reported? (1 point)
36. Have the definitions of confidence intervals been included? (1 point)

Plan evaluations by clinicians

The potential role of clinicians in assessing plan acceptability has been described above. For many planning studies, formal plan evaluations or comparisons by clinicians, e.g. using visual analogue scales [10], can provide important added value regarding the quality of acceptable plans. Clinicians give an overall assessment of the plans, including any trade-offs between planning objectives, high and low dose conformality, etc. Moreover, new techniques or planning approaches will only be introduced clinically if they are preferred by treating clinicians. For plan comparisons, blinded clinician scoring for avoiding bias is recommended [10].

37. Have clinicians scored plans to assess quality? (1 point)
38. Were plan comparisons by clinicians blinded? (1 point)

Predicted tumour control probability and normal tissue complication probabilities for plan evaluation and comparison

The potential clinical impact of plans can sometimes be estimated using predicted tumour control probabilities (TCP) and/or predicted normal tissue complication probabilities (NTCP). However, the underlying models may have large uncertainties. Therefore, TCPs and NTCPs can generally only be used to complement other reporting, e.g. on obtained DVH parameters.

39. Have any applied TCP models been described and referenced? (1 point)
40. Have any applied NTCP models been described and referenced? (1 point)

Plan deliverability and complexity

Depending on the study aim, it may be important to verify that generated plans can indeed be delivered with sufficient accuracy. The most direct approach for this is by dosimetric measurements with an appropriate detector, during plan delivery at a treatment unit. Often the dosimetric analyses are performed with gamma

index analysis [16]. These should be specified in detail, including specification of local or global gamma index, absolute or relative comparison, dose difference and distance criteria, dose difference normalisation point, dose low gradient region, low dose cut-off [17,18]. The standard and experimental plans should preferentially be measured in pairs, so that differences in measurement conditions have minimal impact on the two measurements.

In plan comparisons, sometimes plan parameters such as MU, mean leaf distance, etc. are used to demonstrate that plans generated with a novel planning technique are not more complex than those from another conventional planning approach. However, there is a lack of agreement on the general applicability of proposed plan complexity parameters in terms of prediction of plan deliverability [19].

41. Have methods used to assess plan deliverability and complexity been described in sufficient detail? (1 point)

Composite plan quality metrics

Composite plan quality metrics, used in commercial products such as PlanIQ, Mobius, etc. are sometimes used in planning studies. However, these composite metrics should be clearly motivated for the specific study (e.g. based on literature), and it should be clearly specified how they are calculated, considering all planning aims with their priorities [20,21]. Typically, composite plan quality metrics should be reported in addition to dose parameters rather than instead.

42. Is there a sufficient basis (e.g. in the literature) for any selected composite plan quality metrics? (1 point)
43. Is there an adequate description of the calculation of the composite plan quality metrics? (1 point)

Planning and delivery times

Efficiency in treatment planning and treatment delivery is often of high interest in planning studies, complementing the information on acceptability, dosimetric plan quality and deliverability. For treatment planning, a clear distinction between hands-on planning time and calculation time may be needed. It is important to detail the applied hardware used. Delivery times can be defined in various ways, e.g. total beam-on time, total time to deliver the dose including gantry and/or couch motions, etc. The most direct way to establish these is by measurement at the treatment unit. Sometimes delivery times may also be estimated by dedicated prediction algorithms.

44. Has measurement of planning times been described in sufficient detail? (1 point)
45. Has the establishment of delivery times been described in sufficient detail? (1 point)

Statistical analysis

In planning studies, often (paired) differences in plan parameters are gathered for plans generated with different planning approaches (e.g. manual vs. automatic planning), or with different treatment approaches (e.g. photons vs. protons). Depending on the planning study scope and on the approach to patient selection, statistical analysis of results may or may not be applicable. Avoiding meaningless statistics is as important as providing an appropriate statistical analysis where needed. When patients were hand-selected, for example, to demonstrate the advantages of a novel planning technique for patients with specific characteristics, statistical analysis is usually not applicable. When a planning study aims to conclude on the average performance of planning techniques,

patient selection may have to be random and appropriate statistical testing needs to be performed to assess the statistical significance of the results obtained in terms of p-values, confidence intervals, etc.

In case multiple tests are performed to answer a research question, corrections might be applicable to decrease the risk of false conclusions, i.e. correct for the number of tests or compared scenarios. As explained above in 'Patient cohort', the number of patients included in a planning study needs to be explained and justified. Especially, in case no statistically significant differences are found between planning approaches or treatment techniques, it can be informative to calculate the smallest difference which could be resolved.

46. Have proper statistical methods been used and described in sufficient detail? (5 points, mandatory)
47. In case of multiple testing for research questions, has this been handled appropriately? (1 point)

Considerations for reporting the results: the results section

The study produces a range of results that need to be structured, analysed and evaluated. The reported results should provide sufficient data to answer the posed research questions, leading to the study's findings and conclusions. Generally, only a summary of all the generated data is presented (presenting information instead of data). The raw data may then be included in supplementary material, which can provide useful information for detailed analyses by readers. The results should be presented in a structured coherent way, together, and separate from any discussion or subjective observations.

Generation of appropriate tables and figures is of crucial importance; they should provide a clear graphical representation of the results obtained, ideally in a format that conclusions are immediately clear. However, the figures should not usually be a repetition of the tables, or vice versa. In the recommendations for the 'Materials and Methods' section, several ways for describing differences in dosimetric plan parameters, including dose-volume parameters, population-mean DVHs, and TCPs and NTCPs are described.

48. Does the provided data contribute to (at least partly) answering all aspects of the research questions, e.g. plan acceptability, dosimetric quality, deliverability and planning and delivery times? (10 points, mandatory)

Dose distribution reporting

Although there should be a focus on the research questions, it is important that there are complete summaries of the characteristics of the generated dose distributions; in the end, full plans need to be considered, not only some aspects. For example, if the research question is about high doses, also information on low doses, conformality, etc. as observed in the patient cohort should be provided. If the research question is on OAR doses, also PTV doses should be reported, etc. Study-specific dose metrics may have to be supplemented with more common dose metrics, so comparisons can be made with other available reported work. It is often useful to add data (e.g. dose distributions, DVHs) for an example patient.

49. Are complete summaries of the dose distributions in the patient cohort provided (low doses, high doses, OARs, PTV, patient, etc.)? (5 points, mandatory)
50. Are tables and figures optimised to clearly present the results obtained? (1 point)

51. Have the answers to the research questions been illustrated for an example patient by providing dose distributions, DVHs, etc.? (1 point)

Plan acceptability reporting – minor and major protocol deviations

Plan acceptability should state whether the plan fulfils relevant protocol requirements, and which parameters are not acceptable if any. If deviations were acceptable in the study design, this should be stated in the methods and reported in the results, including any minor and major protocol deviations.

52. In case of treatment technique or planning technique comparisons, was plan acceptability reported separately for each technique? (1 point)
53. Has plan acceptability been reported in sufficient detail: how many plans were acceptable, how many were not and for what reasons (e.g. violation of hard constraints, violation of soft constraints, other reasons)? (1 point)
54. Was there adequate reporting of minor and major protocol deviations? (1 point)

Deliverability and complexity reporting

Where deliverability QA measurements are performed, results should be evaluated with the criteria used in clinical routine. However, apart from answers of acceptable/not acceptable, numerical test results should also be provided (e.g. obtained gamma passing rates, etc.), since the clinical thresholds are only a bare minimum. Along with the QA measurements, or as an alternative, complexity parameters like MU, number of segments, mean leaf separation, etc. may be reported [17].

55. Has the deliverability of the plans been adequately reported? (1 point)
56. Have plan deliverability and complexity been investigated in sufficient detail in relation to the posed research questions? (1 point)

Planning and delivery times reporting

Depending on the study it may be highly relevant to report on planning and delivery times. If planning or delivery takes too long, the clinical application may not be feasible. Differences in planning times for various techniques may indicate a study bias

57. Have planning and delivery times been adequately evaluated and reported? (1 point)

Patient-specific analyses reporting

Planning studies often compare groups of paired plans (e.g. for each patient a photon and a proton plan) to investigate population-mean differences with their statistical significance. In addition, enough detail should be provided for individual patients. For example, if there is no statistically significant difference for a plan parameter, there may be clinically significant differences for individual patients. These should be explicitly described, as they may be highly relevant for sub-groups or increasing personalised medicine considerations.

It is also important to sufficiently report results for so-called outlier patients. If they are excluded from population analyses this needs to be well motivated and explained.

58. Is there sufficient description of inter-patient variations in the results presented? (1 point)

59. Have outlier patients been reported and has any exclusion from population analyses been sufficiently motivated and explained? (1 point)

Statistical reporting

If applicable, the statistical reporting for the primary research question should be clear and stand out, compared to the other secondary research questions. It is common practice to show one significant digit of p-values. Any p-value below 0.001 could be reported as <0.001. Avoid reporting of p-values above the significant threshold as non-significant or NS, i.e. report the numbers. In addition to p-values, it is advised to also report confidence intervals or equivalent where available since this can be more informative than p-values.

60. Are the p-values reported appropriately? (1 point)
61. Are there confidence intervals for the appropriate parameters? (1 point)

Considerations for the interpretation and discussion of the results: the discussion section

The Discussion section often starts with an overall interpretation by the scientists of the presented results in terms of answers to the research questions, posed at the initial design stage, as laid out in the Introduction. After that, the interpretation is put in context (e.g. of other literature) and discussed in terms of limitations, clinical significance, clinical applicability and future work.

62. Is there an overall interpretation of the data presented in the Results section as to how the posed research questions are answered? (10 points, mandatory)

Comparison with literature

Mostly, planning studies are performed where there are some existing knowledge or knowledge gaps reported in the literature. There may also be publications describing answers to the posed research questions using other methodology. The obtained answers to the research questions should be discussed in the context of the relevant literature.

63. Has the study been sufficiently discussed in the context of existing literature? (5 points, mandatory)

Clinical and statistical significance

Generally, discussion of the results focusses on statistically significant results. However, statistical significance does not mean that observed differences are clinically meaningful, e.g. in the case that they are small. Therefore, beyond statistical significance, the study results also should be discussed in the context of clinical significance. This can be done, for example, by referring to literature results describing the clinical impact of observed differences.

64. Does the discussion focus on statistically significant results? (1 point)
65. Is the potential clinical significance of the results clearly discussed (assuming practical application would be feasible)? (5 points, mandatory)

Clinical applicability of the study results

Clinical applicability can be an important point of discussion. A description of any limitations (e.g. for selected patients only) or

restrictions (e.g. availability of certain equipment) may be of high value.

66. Is future clinical applicability sufficiently discussed? (1 point)

Study limitations

Most studies have some limitations in the methods of obtaining answers and/or in the provided answers to the posed research questions. These should be clearly addressed in the Discussion section. A major limitation may be bias, as described in 'Considerations for the methodology: the Materials and Methods section'.

67. Has the impact of the study limitations on the provided answers to the research questions been sufficiently discussed? (10 points, mandatory)

Future work

In many studies, a description of future work can be highly relevant. This can relate to new ideas for new studies, including those that could possibly avoid current study limitations, plans for clinical application of the study results, research for different tumour sites, etc.

68. Has the potential future work arising from the study been discussed? (1 point)

Considerations for the conclusion from the planning study: the Conclusion section

The conclusion from any such study should focus on the answers to the primary research question. Accurate descriptions with positive and negative observations are needed. The presented conclusions should be fully supported by the obtained results, without wider conjecture.

69. Do the presented conclusions represent answers to the posed research questions? (5 points, mandatory)
70. Are the conclusions supported by the results? (5 points, mandatory)
71. Are the conclusions a fair summary of all results? (5 points, mandatory)

Considerations for supplementary sections of published planning studies: the Supplementary Materials section

Supplementary materials

The data presented in the Results section of the main paper often only consist of a concise overview of all the (raw) data generated. This can have several reasons, including raw data having been processed to optimally answer the research questions, a specific journal setting a limit on the number of allowed figures and tables, lack of space in the main body for more detail, etc. In these cases, an electronic appendix may be added. It should be noted that this appendix also should meet adequate quality levels of readability, clarity, coherence, etc., with clear links and support for the main report. There should be clear descriptions of the presented data and of the methods used to produce them. Any data presentation should be guided by the FAIR Data Principles [22]. As much as possible, an electronic appendix has to make sense and be readable by itself, i.e. without too much reference to the main paper.

Table 1
RATING score sheet.

RATING score sheet		Points	Applicable/ relevant	Answer yes
Questions for the Introduction				
<i>The study aim formulated by research questions</i>				
1	Does the study have a concise and precise study aim, defined with a restricted number of interconnected questions?	10	<input type="checkbox"/>	<input type="checkbox"/>
<i>The motivation for the research questions</i>				
2	Has relevant up to date literature been included to support the need for the current study?	5	<input type="checkbox"/>	<input type="checkbox"/>
3	Does the study address an existing knowledge gap?	10	<input type="checkbox"/>	<input type="checkbox"/>
Questions for Materials and Methods				
4	Is the global study design adequate for answering the posed research questions?	10	<input type="checkbox"/>	<input type="checkbox"/>
5	Is the global study design described in sufficient detail for others to interpret and reproduce the results?	5	<input type="checkbox"/>	<input type="checkbox"/>
<i>Patient cohort</i>				
6	Are the inclusion and exclusion criteria of the patient cohort described?	1	<input type="checkbox"/>	<input type="checkbox"/>
7	Is the clinical patient information of the cohort presented, including disease type, site(s) and clinical staging?	1	<input type="checkbox"/>	<input type="checkbox"/>
8	Is the included number of patients stated, explained and justified?	1	<input type="checkbox"/>	<input type="checkbox"/>
9	Has there been consideration of the need for ethical and/or legal approval for the study and if needed, is there a statement about this?	5	<input type="checkbox"/>	<input type="checkbox"/>
<i>Imaging procedures</i>				
10	Have the scanning parameters been reported in sufficient detail (image modalities, equipment model, slice thickness, voxel size, patient position (e.g. head first, supine, etc.) etc.)?	1	<input type="checkbox"/>	<input type="checkbox"/>
11	Has the applied immobilisation equipment been described, (e.g. vendor and type, standard settings, etc.) where relevant?	1	<input type="checkbox"/>	<input type="checkbox"/>
<i>Treatment machine and settings</i>				
12	Have the treatment machine and relevant parameters been described with sufficient detail (model, beam energy, MLC, etc.)?	1	<input type="checkbox"/>	<input type="checkbox"/>
13	Have the monitor unit reference conditions been defined, where relevant?	1	<input type="checkbox"/>	<input type="checkbox"/>
<i>Definition of targets and OARs</i>				
14	Has GTV definition been described in sufficient detail, with references if possible?	1	<input type="checkbox"/>	<input type="checkbox"/>
15	Has CTV definition been described in sufficient detail, with references if possible?	1	<input type="checkbox"/>	<input type="checkbox"/>
16	Has the establishment of PTVs (or alternatively robustness settings) been described in sufficient detail?	1	<input type="checkbox"/>	<input type="checkbox"/>
17	Have PTV sizes in the patient cohort been described?	1	<input type="checkbox"/>	<input type="checkbox"/>
18	Have OAR definitions been described in sufficient detail, with references if possible?	1	<input type="checkbox"/>	<input type="checkbox"/>
19	Have PRV margins been described in sufficient detail, with references if available?	1	<input type="checkbox"/>	<input type="checkbox"/>

72. Is the information presented in the supplementary material of sufficient relevance? (1 point)

73. Is the presentation of the included information of sufficient quality, including readability? (1 point)

Sharing data must conform to legal regulations on patient data confidentiality, which may vary in different jurisdictions. However, it is possible to publish anonymised data if the applicable regulations are followed. Whilst DICOM CT images, structures, resulting dose distributions and plans are not commonly published today,

scientists should keep in mind that a planning study's contribution to the scientific community is significantly higher if the underlying data is available. Even where data cannot be made fully publicly available, it may be possible to indicate a willingness to share data with other researchers under an appropriate written agreement. Mostly, it will be possible to share the analysis code.

74. Has sufficient underlying data been made available or a willingness to share data been indicated, within local data sharing restrictions? (5 points, mandatory)

Treatment planning system and dose calculation

20	Have all applied dose calculation algorithms been described in sufficient detail?	1	<input type="checkbox"/>	<input type="checkbox"/>
21	For any commercial software used, have the manufacturer, algorithms and specific versions been stated?	1	<input type="checkbox"/>	<input type="checkbox"/>
22	Have all relevant user parameters and settings in the TPS been reported, e.g. beams, dose grid, control point spacing?	1	<input type="checkbox"/>	<input type="checkbox"/>
23	Have all volumes been evaluated with the same software/methodology?	1	<input type="checkbox"/>	<input type="checkbox"/>

Planning aims and optimisation

24	Are clear planning aims defined, including imposed hard constraints and planning objectives (with or without soft constraints)?	5	<input type="checkbox"/>	<input type="checkbox"/>
25	Has the ranking of planning objectives (priorities) been described?	5	<input type="checkbox"/>	<input type="checkbox"/>
26	Is the dose prescription clearly defined?	10	<input type="checkbox"/>	<input type="checkbox"/>
27	Is there a narrative description of the applied optimisation process, including the handling of all objectives with their ranking?	5	<input type="checkbox"/>	<input type="checkbox"/>
28	If manual intervention during or after optimisation is allowed, has this been described?	1	<input type="checkbox"/>	<input type="checkbox"/>

Bias mitigation

29	Have enough study details been provided such that bias issues could be noted?	5	<input type="checkbox"/>	<input type="checkbox"/>
30	Has bias been sufficiently mitigated to reliably answer the posed research question?	10	<input type="checkbox"/>	<input type="checkbox"/>

Plan acceptability – minor and major protocol deviations

31	Was the procedure for assessment of plan acceptability well-described?	1	<input type="checkbox"/>	<input type="checkbox"/>
32	Was the procedure for assessment of minor and major protocol deviations well described?	1	<input type="checkbox"/>	<input type="checkbox"/>

Plan (re-)normalisation for plan comparisons

33	Has plan (re-)normalisation been described sufficiently?	1	<input type="checkbox"/>	<input type="checkbox"/>
----	--	---	--------------------------	--------------------------

Dose-volume parameters for plan evaluation and comparison

34	Have sufficiently comprehensive dose-volume parameters been used for plan evaluations and comparisons?	5	<input type="checkbox"/>	<input type="checkbox"/>
----	--	---	--------------------------	--------------------------

Population-mean DVHs

35	Has the algorithm for creating population-mean/median DVHs been reported?	1	<input type="checkbox"/>	<input type="checkbox"/>
36	Have the definitions of confidence intervals been included?	1	<input type="checkbox"/>	<input type="checkbox"/>

Plan evaluations by clinicians

37	Have clinicians scored plans to assess quality?	1	<input type="checkbox"/>	<input type="checkbox"/>
38	Were plan comparisons by clinicians blinded?	1	<input type="checkbox"/>	<input type="checkbox"/>

*Predicted tumour control probability and normal tissue**complication probabilities for plan evaluation and comparison*

39	Have any applied TCP models been described and referenced?	1	<input type="checkbox"/>	<input type="checkbox"/>
40	Have any applied NTCP models been described and referenced?	1	<input type="checkbox"/>	<input type="checkbox"/>

Plan deliverability and complexity

41	Have methods used to assess plan deliverability and complexity been described in sufficient detail?	1	<input type="checkbox"/>	<input type="checkbox"/>
----	---	---	--------------------------	--------------------------

Composite plan quality metrics

42	Is there a sufficient basis (e.g. in the literature) for any selected composite plan quality metrics?	1	<input type="checkbox"/>	<input type="checkbox"/>
----	---	---	--------------------------	--------------------------

(continued on next page)

RATING score sheet

All questions above are collected in the RATING score sheet and given weights reflecting their relative importance (Table 1 and Excel spreadsheet in the electronic appendix). The weights are indicative and constrained to a limited set. Questions can be selected as Applicable (or Not) to the study, although some are defined as applicable to all studies, and then answered as Yes, or left blank if No. The RATING score is the normalised weighted-sum of all mandatory (blue shading) and applicable 'Yes' answers,

having a maximum value of 100%. In the preparation phase of a planning study, by examining high-weight or insufficiently-considered issues in the list, the score sheet can assist in ensuring a high-quality research question, study design and setup; and in the writing phase of the study, a high-quality report.

75. Is the RATING score added to the manuscript? (5 points, mandatory)
76. Is the accompanying question table added to the cover letter or the supplementary material? (1 point)

43	Is there an adequate description of the calculation of the composite plan quality metrics?	1	<input type="checkbox"/>	<input type="checkbox"/>
<i>Planning and delivery times</i>				
44	Has measurement of planning times been described in sufficient detail?	1	<input type="checkbox"/>	<input type="checkbox"/>
45	Has the establishment of delivery times been described in sufficient detail?	1	<input type="checkbox"/>	<input type="checkbox"/>
<i>Statistical analysis</i>				
46	Have proper statistical methods been used and described in sufficient detail?	5	<input checked="" type="checkbox"/>	<input type="checkbox"/>
47	In case of multiple testing for research questions, has this been handled appropriately?	1	<input type="checkbox"/>	<input type="checkbox"/>
Questions for Results				
48	Does the provided data contribute to (at least partly) answering all aspects of the research questions, e.g. plan acceptability, dosimetric quality, deliverability and planning and delivery times?	10	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<i>Dose distribution reporting</i>				
49	Are complete summaries of the dose distributions in the patient cohort provided (low doses, high doses, OARs, PTV, patient, etc.)?	5	<input checked="" type="checkbox"/>	<input type="checkbox"/>
50	Are tables and figures optimised to clearly present the results obtained?	1	<input type="checkbox"/>	<input type="checkbox"/>
51	Have the answers to the research questions been illustrated for an example patient by providing dose distributions, DVHs, etc.?	1	<input type="checkbox"/>	<input type="checkbox"/>
<i>Plan acceptability reporting – minor and major protocol deviations</i>				
52	In case of treatment technique or planning technique comparisons, was plan acceptability reported separately for each technique?	1	<input type="checkbox"/>	<input type="checkbox"/>
53	Has plan acceptability been reported in sufficient detail: how many plans were acceptable, how many were not and for what reasons (e.g. violation of hard constraints, violation of soft constraints, other reasons)?	1	<input type="checkbox"/>	<input type="checkbox"/>
54	Was there adequate reporting of minor and major protocol deviations?	1	<input type="checkbox"/>	<input type="checkbox"/>
<i>Deliverability and complexity reporting</i>				
55	Has the deliverability of the plans been adequately reported?	1	<input type="checkbox"/>	<input type="checkbox"/>
56	Have plan deliverability and complexity been investigated in sufficient detail in relation to the posed research questions?	1	<input type="checkbox"/>	<input type="checkbox"/>
<i>Planning and delivery times reporting</i>				
57	Have planning and delivery times been adequately evaluated and reported?	1	<input type="checkbox"/>	<input type="checkbox"/>
<i>Patient-specific analyses reporting</i>				
58	Is there sufficient description of inter-patient variations in the results presented?	1	<input type="checkbox"/>	<input type="checkbox"/>
59	Have outlier patients been reported and has any exclusion from population analyses been sufficiently motivated and explained?	1	<input type="checkbox"/>	<input type="checkbox"/>
<i>Statistical reporting</i>				
60	Are the p-values reported appropriately?	1	<input type="checkbox"/>	<input type="checkbox"/>
61	Are there confidence intervals for the appropriate parameters?	1	<input type="checkbox"/>	<input type="checkbox"/>

Summary

This 'RATING' framework is proposed with the aim of improving the scientific quality of treatment planning studies and papers reporting them. It can be used at any stage, from the initial planning phase to inform the study design, to the end phase to assist in study evaluation and reporting. It is hoped it will be a

useful and consistent tool for researchers and for reviewers and editors.

Conflict of Interest Statement

There are no conflict of interest in the author group in the current RATING project.

Questions for discussions				
62	Is there an overall interpretation of the data presented in the Results section as to how the posed research questions are answered?	10		<input type="checkbox"/>
<i>Comparison with literature</i>				
63	Has the study been sufficiently discussed in the context of existing literature?	5		<input type="checkbox"/>
<i>Clinical and statistical significance</i>				
64	Does the discussion focus on statistically significant results?	1	<input type="checkbox"/>	<input type="checkbox"/>
65	Is the potential clinical significance of the results clearly discussed (assuming practical application would be feasible)?	5		<input type="checkbox"/>
<i>Clinical applicability of the study</i>				
66	Is future the clinical applicability sufficiently discussed?	1	<input type="checkbox"/>	<input type="checkbox"/>
<i>Study limitations</i>				
67	Has the impact of the study limitations on the provided answers to the research questions been sufficiently discussed?	10		<input type="checkbox"/>
<i>Future work</i>				
68	Has the potential future work arising from the study been discussed?	1	<input type="checkbox"/>	<input type="checkbox"/>
Questions for conclusions				
69	Do the presented conclusions represent answers to the posed research questions?	5		<input type="checkbox"/>
70	Are the conclusions fully supported by the results?	5	<input type="checkbox"/>	<input type="checkbox"/>
71	Are the conclusions a fair summary of all results?	5	<input type="checkbox"/>	<input type="checkbox"/>
Questions for supplementary				
<i>Supplementary materials</i>				
72	Is the information presented in the supplementary material of sufficient relevance?	1	<input type="checkbox"/>	<input type="checkbox"/>
73	Is the presentation of the included information of sufficient quality, including readability?	1	<input type="checkbox"/>	<input type="checkbox"/>
74	Has sufficient underlying data been made available or a willingness to share data been indicated, within local data sharing restrictions?	5		<input type="checkbox"/>
RATING remarks				
75	Is the RATING score added to the manuscript?	5		<input type="checkbox"/>
76	Is the accompanying question table added to the cover letter or the supplementary material?	1	<input type="checkbox"/>	<input type="checkbox"/>

All questions from the text are listed. And each is assigned a weight. Most questions can be selected as Relevant/Applicable for the study or not, although some are considered relevant for all planning studies (marked blue). Questions answers can be Yes (affirmed with a 'tick' in the table) or No (leave box blank). From the answers to the Relevant/Applicable questions, a normalised weighted-sum score, the RATING score, is calculated with a maximum of 100%. The spreadsheet in the supplementary material automates this. The weights assigned to the questions are indicative. They represent the author group consensus on relative importance but constrained to a limited set of levels. Different researchers might suggest different specific weights for particular questions in a given context; however, the aim is to have a consistent approach for scoring without too many levels.

Acknowledgements

Dirk Verellen is acknowledged for his role as co-chair of the Automated Planning workshop at the 1st ESTRO physics workshop in Glasgow in 2017, which initiated the RATING project. CRH acknowledge support from DCCC Radiotherapy - The Danish National Research Center for Radiotherapy, Danish Cancer Society and Danish Comprehensive Cancer Center. MH acknowledge funding from the UK National Measurement System.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.radonc.2020.09.033>.

References

- [1] Bentley RE, Milan J. An interactive digital computer system for radiotherapy treatment planning. *Br J Radiol* 1971;44:826–33.
- [2] Garibaldi C, Jerezek-Fossa BA, Marvaso G, Dicuonzo S, Rojas DP, Cattani F, et al. Recent advances in radiation oncology. *e-Cancer Med Sci* 2017;11:785.
- [3] Brahme A. Optimization of stationary and moving beam radiation therapy techniques. *Radiother Oncol* 1988;12:129–40.
- [4] Otto K. Volumetric modulated arc therapy: IMRT in a single gantry arc. *Med Phys* 2008;35:310–7.
- [5] Bertelsen A, Hansen CR, Johansen J, Brink C. Single arc volumetric modulated arc therapy of head and neck cancer. *Radiother Oncol* 2010;95:142–8.
- [6] Unkelbach J, Paganetti H. Robust proton treatment planning: physical and biological optimization. *Semin Radiat Oncol*. 2018;28:88–96.
- [7] Hussein M, Heijmen BJM, Verellen D, Nisbet A. Automation in intensity modulated radiotherapy treatment planning—a review of recent innovations. *Br J Radiol* 2018;91:20180270.
- [8] Cozzi L, Heijmen BJM, Muren LP. Advanced treatment planning strategies to enhance quality and efficiency of radiotherapy. *Physics and Imaging in Radiation Oncology*. 2019;11:69–70.
- [9] Sonke JJ, Aznar M, Rasch C. Adaptive radiotherapy for anatomical changes. *Semin Radiat Oncol*. 2019;29:245–57.
- [10] Gurney-Champion OJ, Mahmood F, van Schie M, Julian R, George B, Philippens MEP, et al. Quantitative imaging for radiotherapy purposes. *Radiother Oncol* 2020;146:66–75.
- [11] Nystrom H, Jensen MF, Nystrom PW. Treatment planning for proton therapy: what is needed in the next 10 years?. *Br J Radiol* 2020;93:20190304.
- [12] Yartsev S, Muren LP, Thwaites DI. Treatment planning studies in radiotherapy. *Radiother Oncol* 2013;109:342–3.
- [13] Winsberg E. Computer simulations in science. Winter 2019 ed.; Metaphysics Research Lab, Stanford University; 2019.
- [14] Kaizer JS, Heller AK, Oberkampf WL. Scientific computer simulation review. *Reliab Eng Syst Saf* 2015;138:210–8.
- [15] Ebert MA, Haworth A, Kearvell R, Hooton B, Hug B, Spry NA, et al. Comparison of DVH data from multiple radiotherapy treatment planning systems. *Phys Med Biol* 2010;55:N337–46.

Radiotherapy Treatment planning study Guidelines (RATING)

- [16] Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. *Med Phys* 1998;25:656–61.
- [17] Miften M, Olch A, Mihailidis D, Moran J, Pawlicki T, Molineu A, et al. Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218. *Med Phys* 2018;45:e53–83.
- [18] Pogson EM, Aruguman S, Hansen CR, Currie M, Oborn BM, Blake SJ, et al. Multi-institutional comparison of simulated treatment delivery errors in sslMRT, manually planned VMAT and autoplan-VMAT plans for nasopharyngeal radiotherapy. *Phys Med* 2017;42:55–66.
- [19] Kamperis E, Kodona C, Hatziiioannou K, Giannouzakos V. Complexity in radiation therapy: it's complicated. *Int J Radiat Oncol Biol Phys* 2020;106:182–4.
- [20] Chiavassa S, Bessieres I, Edouard M, Mathot M, Moignier A. Complexity metrics for IMRT and VMAT plans: a review of current literature and applications. *Br J Radiol* 2019;92:20190270.
- [21] Glenn MC, Hernandez V, Saez J, Followill DS, Howell RM, Pollard-Larkin JM, et al. Treatment plan complexity does not predict IROC Houston anthropomorphic head and neck phantom performance. *Phys Med Biol* 2018;63.
- [22] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016;3.